9-2018

# Marikina flood hazard models using historical data of water level

Elvira P. De Lara-Tuprio
*Ateneo de Manila University*

Evangeline P. Bautista
*Ateneo de Manila University*

Reginaldo M. Marcelo
*Ateneo de Manila University*

Ramil T. Bataller
*Ateneo de Manila University*

Divino Angelo B. Esteban

*See next page for additional authors*



## Recommended Citation

De Lara-Tuprio, E.P., Bautista, E.P., Marcelo, R.M., Bataller, R.T., Esteban, D.A.B., Yutuc, Y.P.B. (2018/09). Marikina flood hazard models using historical data of water level. Philippine Journal of Science, 147(3), 373-382.

## Authors
Elvira P. De Lara-Tuprio, Evangeline P. Bautista, Reginaldo M. Marcelo, Ramil T. Bataller, Divino Angelo B. Esteban, and Yvanne Paolo B. Yutuc

**Philippine Journal of Science**
147 (3): 373-382, September 2018
ISSN 0031 - 7683
Date Received: 02 Aug 2017

# Marikina Flood Hazard Models Using Historical Data of Water Level

**Elvira P. de Lara-Tuprio*, Evangeline P. Bautista, Reginaldo M. Marcelo, Ramil T. Bataller, Divino Angelo B. Esteban, and Yvanne Paolo B. Yutuc**

Mathematics Department, School of Science and Engineering,
Ateneo de Manila University, Loyola Heights, Quezon City 1108 Philippines

**In this paper, ten-year historical data of water levels recorded at Sto. Niño, Marikina station of MMDA-EFCOS were analysed and processed to determine the number of times per year (annual frequency) that critical levels of the Marikina River near the Sto. Niño station were reached and for how long (duration). Probability distributions for the annual frequency and duration were then fitted to the samples obtained. Monte Carlo simulation was applied in order to generate possible realizations of the random variables. Summary statistics were then obtained from the simulated values. Finally, backtesting using historical data of water levels after the period of model development was performed to check the validity of the models. The results showed that the models obtained were reliable. The results of this study may be used to guide the local government of Marikina in planning the needed resources in order to sufficiently respond in times of flooding incidents.**

Key words: annual frequency, flood, Monte Carlo simulation, probability distribution

## INTRODUCTION

Floods brought about by typhoons and thunderstorms are a perennial problem in Metro Manila, especially in low-lying areas like Marikina City. The problem may be considered as a combination of natural and manmade disasters. Damage to private and government properties, loss of lives, loss of livelihood, and disruption in social and economic activities are just some of the negative effects of flooding.

Marikina City is located in the eastern border of Metro Manila and bound on the west by Quezon City, on the south by Pasig and Cainta, on the east by Antipolo City, and on the north by San Mateo, Rizal. Marikina lies on what is known as the Marikina Valley, with the Sierra Madre to the east and the Quezon City hills to the east (Iglesias 2008). The Marikina River cuts across the mid-west portion of the city, while its tributary is on the north flowing slightly through San Mateo.

Being a valley with a river running through it, Marikina has had a long history of problems with flooding. The river swells and overflows in times of excessive rain and at the same time, the city becomes a catch basin of water coming from Quezon City and Rizal. The flood of 1992 – which inundated 27.52% of the city's land area and affected 10,000 households – is a landmark as far as the city is concerned (Francisco 2015). The local government – headed by then Marikina Mayor Bayani Fernando – the Department of Public Works and Highways (DPWH), and Metropolitan Manila Development Authority (MMDA) worked together to implement flood control projects in the city. The local government also created several ordinances and resolutions addressing the city's disaster preparedness, mitigation, and response (Yu & Sajor 2008). The progress, however, took a setback in Sep 2009 when typhoon Ondoy brought in a record 448.5 mm of rain in just a span of 12 h, flooding the central part of Luzon

Corresponding author: edelaratuprio@ateneo.edu

and causing a drastic 10.99 m rise in the water level of Marikina River (NOAH 2009).

Ondoy wreaked havoc on Marikina City as it was the one most affected by the disaster. It showed that despite all the progress, Marikina is still not prepared for calamities of this magnitude. From the paper of Sato and Nakasu (2011), "The large depth of flood water combined with the sharp increase in water level resulted in a severe inundation that spread across the entire flood plain. Floodwater depths of 7 m were recorded in some places although the flooding was short-lived. The flood caused damage on a large scale, affecting many people and causing 121 deaths (half the total for Metro Manila). The flood plain lies mostly in Marikina City, but part of it extends upstream into Quezon City. In Marikina City alone, there were approximately 180,000 people affected by the flooding."

Post-Ondoy, several more incidents of heavy flooding in Marikina occurred. Included here are the monsoon rains in Aug 2012 that caused the Marikina River to swell to at least 19 m, and the 2014 flooding caused again by the monsoon rains and worsened by typhoon Mario.

This study aims to describe the annual frequency of flooding events in Marikina City – particularly in areas near the Marikina River – together with the duration in hours of each event based on historical data of hourly record of water level obtained from Effective Flood Control Operation System (EFCOS) office, which is under the management of MMDA. The data obtained were then processed to determine appropriate probability distributions that can be used to assess the likelihood of future occurrences of flooding.

The approach used in this study differs from studies dealing with flood hazard models that are derived from geomorphological and hydrological-hydraulic methods (Abon et al. 2011; Apel et al. 2006; Badilla 2008; Cezar 2008). This study will use neither of these methods and instead focus on the information that can be extracted from historical data of water level and apply probabilistic models. The idea is derived from the works of Born & Martin (2006), Cardona and co-authors (2003), Grossi & Kunreuther (2005), Kundu and co-authors (2012), Mechler & The Risk to Resilience Study Team (2008), Mechler (2005), and Michel-Kerjan (2013), The proposed method is not meant to replace but rather to supplement existing approaches. Note that – since the basis of the proposed approach is historical data, then when used for forecasting future flood frequency and duration – the implicit assumption is that there are no interventions done yet by local government to mitigate future flood hazard over the period of measurement.

This paper is organized as follows. Section I describes the area of study, Marikina City, and the objective of this study. Section II describes the raw and processed data and the procedure taken. Section III presents the results and discussion of the model development. Section IV validates the models obtained. Finally, section V gives a summary, conclusion, and some recommendations.

## DATA AND METHODOLOGY

The main data used in this study are the water level measurements taken at the Sto. Niño station of EFCOS. These measurements have been the basis of the Marikina Disaster Risk Reduction and Management Office (MDRRMO) to monitor the flooding in the city and to warn the residents of barangays near the river of possible flooding, and to initiate evacuation if necessary. Barangays that are usually affected are Nangka, Tumana, Malanday, Tañong, and Sta. Elena (Figure 1).

The alert warnings according to the indicated water level at Sto. Niño are as follows:

Alert Level 1: 15 m to 15.99 m (or less than 16 m), Alert

Alert Level 2: 16 m to 16.99 m, Prepare to evacuate

Alert Level 3: 17 m to 17.99 m, Evacuate to designated centers

Alert Level 4: 18 m and above, Forced evacuation

This study uses the hourly measurements of the water level at Sto. Niño obtained from MMDA from 1 Jan 2002 to 31 Dec 2012 for the model development. For illustration, the water level measurements during the years 2002-2003 are shown in Figure 2.

The entire plot of raw data from 2002 to 2012 is shown in Figure 3. Notice the break in the graph between 2009 and 2010. This set of missing data will be explained further in the succeeding discussions.

The available data consist of 83,558 hourly measurements during the identified period of study. From the expected total of 96,432, there are 12,874 that are missing, mostly during the years 2009 to 2010. See Table 1 for the number of missing measurements per month in each year. A major system failure at the Sto. Niño station caused by Ondoy, which could not be restored until the year later, explains the big number of missing data during that period.

Instead of trying to fill in the about 9,000 missing measurements from 1 Oct 2009 to 30 Sept 2010, the available data from 1 Jan to 30 Sept 2009 and the data from 1 Oct to 31 Dec 2010 were joined to form one year, herein referred to as year 2009-10. Other missing data during the months of Sept 2009 and Oct 2010 were filled based on


**Figure 1.** Map of barangays near Marikina River.

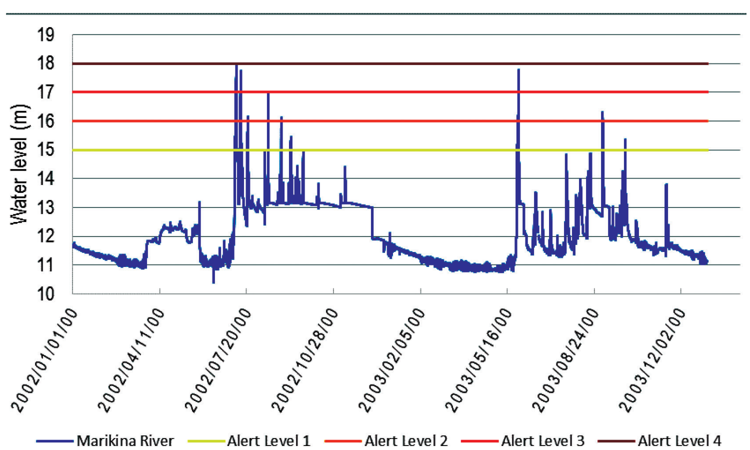


**Figure 2.** Hourly measurements of the water level at the Sto. Niño station of EFCOS project from 1 Jan 2002 to 31 Dec 2003.

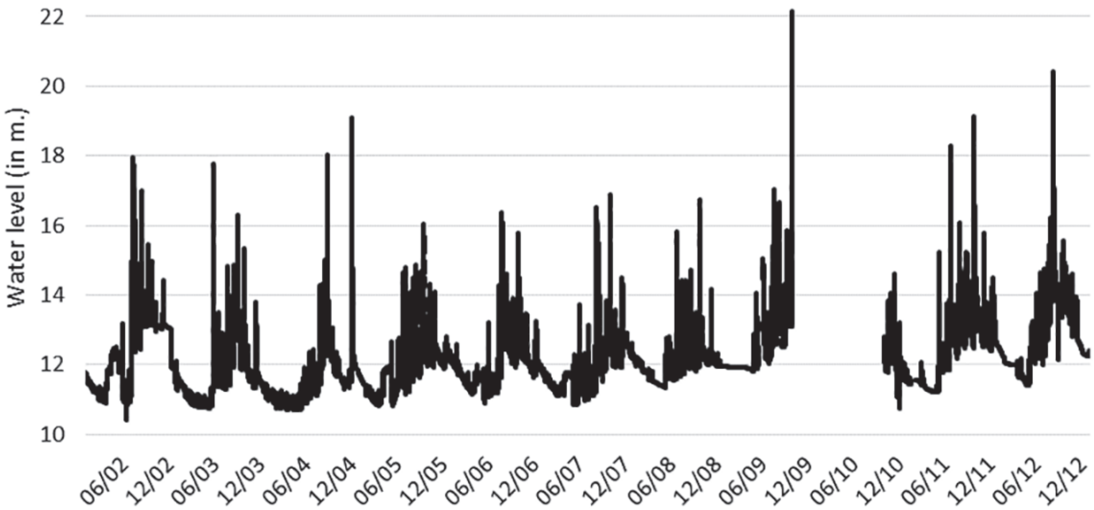


**Figure 3.** Plot of hourly measurements of the water level at the Sto. Niño station of EFCOS project from 1 Jan 2002 to 31 Dec 2012.

**Table 1.** Number of missing hourly measurements of the water level at the Sto. Niño station of the EFCOS project from 01 Jan 2002 to 31 Dec 2012.

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | 0 | 0 | 0 | 0 | 79 | 1 | 1 | 0 | 744 | 513 | 2 |
| Feb | 0 | 0 | 0 | 0 | 78 | 6 | 0 | 0 | 672 | 96 | 0 |
| Mar | 0 | 0 | 0 | 0 | 44 | 8 | 0 | 0 | 744 | 0 | 179 |
| Apr | 0 | 0 | 0 | 37 | 47 | 0 | 0 | 0 | 720 | 0 | 116 |
| May | 0 | 0 | 0 | 0 | 59 | 0 | 0 | 338 | 744 | 50 | 0 |
| Jun | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 194 | 720 | 0 | 0 |
| Jul | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 438 | 744 | 0 | 0 |
| Aug | 0 | 0 | 0 | 9 | 51 | 2 | 24 | 262 | 744 | 0 | 2 |
| Sep | 0 | 0 | 0 | 106 | 60 | 2 | 1 | 318 | 641 | 0 | 8 |
| Oct | 0 | 0 | 0 | 108 | 5 | 3 | 24 | 744 | 208 | 0 | 1 |
| Nov | 0 | 0 | 0 | 170 | 0 | 1 | 0 | 720 | 186 | 5 | 106 |
| Dec | 0 | 0 | 0 | 166 | 0 | 1 | 0 | 744 | 30 | 8 | 29 |
| Total | 0 | 11 | 0 | 596 | 423 | 24 | 50 | 3758 | 6897 | 672 | 443 |
| No. of hours covered | 8760 | 8760 | 8784 | 8760 | 8760 | 8760 | 8784 | 8760 | 8760 | 8760 | 8784 |

news reports on Ondoy and Pepeng, while the rest of the missing data were filled using linear interpolation. As a result, there were a total of 87,672 hourly measurements of the water level at Sto. Niño covering an equivalent of 10 yrs. For illustration, Figure 4 shows the complete data for the combined year 2009-10, from Jan 1 to Dec 31.

From the final set of data of hourly water levels, the annual frequency of water level reaching a particular Alert Level and the length of time per instance that it stayed on or above the lower boundary of that Alert Level were determined. The lower boundaries are 15 m, 16 m, 17 m, and 18 m. Tables 2 to 5 show the results.

The table entries corresponding to Duration constitute sample values of a continuous random variable representing the length of time in hours that water level stays at or above the specified lower boundary, regardless of year. The numbers found at the bottom are considered as values of the random variable representing the annual frequency that such lower boundary is reached. The distributions of these random variables will be derived using statistical tools.

**Table 2.** Number of incidents of Alert Level 1 or higher and length of time in hours per incident from 2002 to 2012.

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009-10 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|
| Duration (h) | 20 | 10 | 1 | 16 | 25 | 27 | 11 | 2 | 8 | 10 |
| | 1 | 29 | 50 | 11 | 3 | 9 | 12 | 6 | 53 | 7 |
| | 27 | 19 | 13 | | 11 | 13 | | 34 | 17 | 5 |
| | 4 | 5 | 15 | | | | | 52 | 9 | 16 |
| | 25 | | | | | | | 16 | 1 | 30 |
| | 7 | | | | | | | 92 | 3 | 2 |
| | 19 | | | | | | | 29 | 26 | 7 |
| | 12 | | | | | | | | 12 | 10 |
| | 8 | | | | | | | | 7 | 18 |
| | | | | | | | | | | 86 |
| | | | | | | | | | | 17 |
| | | | | | | | | | | 4 |
| | | | | | | | | | | 3 |
| | | | | | | | | | | 9 |
| | | | | | | | | | | 8 |
| Number of incidents | 9 | 4 | 4 | 2 | 3 | 3 | 2 | 7 | 9 | 15 |

**Table 3**. Number of incidents of Alert Level 2 or higher and length of time in hours per incident from 2002 to 2012.

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009-10 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|
| Duration (h) | 18 | 18 | 41 | 1 | 10 | 11 | 7 | 17 | 42 | 5 |
| | 17 | 5 | 10 | | | 7 | | 15 | 2 | 4 |
| | 3 | | 8 | | | | | 7 | 18 | 9 |
| | 7 | | | | | | | 81 | 7 | 73 |
| | | | | | | | | 22 | | 10 |
| Number of incidents | 4 | 2 | 3 | 1 | 1 | 2 | 1 | 5 | 4 | 5 |

**Table 4.** Number of incidents of Alert Level 3 or higher and length of time in hours per incident from 2002 to 2012.

| | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009-10 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|
| Duration (h) | 12 | 7 | 11 | | | | | 1 | 17 | 64 |
| | 11 | | 7 | | | | | 69 | 15 | 3 |
| | | | 3 | | | | | 16 | | |
| Number of incidents | 2 | 1 | 3 | 0 | 0 | 0 | 0 | 3 | 2 | 2 |

**Table 5.** Number of incidents of Alert Level 4 and length of time in hours per incident from 2002 to 2012.

| Duration (h) | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009-10 | 2011 | 2012 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2 | | | | | 56 | 5 | 32 |
| | | | 6 | | | | | | 11 | 21 |
| Number of incidents | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 2 |

## RESULTS AND DISCUSSION

There are several statistical tools that can be used to find the probability distribution that best fits a given sample, some of which are Kolmogorov-Smirnov test, Anderson-Darling test, and Chi-squared test. For this study, the Anderson-Darling test was applied. Parsimony was also considered in choosing the distributions. The distributions obtained can then be used to simulate possible values in the future.

Two random variables were considered: discrete random variable for the annual frequency of at least one-hour stay in a water level of specified lower boundary, and continuous distribution for the duration in hours for each occurrence. In the literature, common discrete distributions are Poisson, Geometric, Binomial, and Negative Binomial (Panjer 2006). For continuous distributions, there are several available with number of parameters ranging from one, two, and so on. The following are the distributions that best fit the samples based on Anderson-Darling test. Note that these distributions were obtained separately.

**Table 6**. Distributions for the annual frequency and duration.

| | Annual Frequency | Duration |
|---|---|---|
| Water level ≥15 | Negative binomial n=3 p=0.36802 | Dagum k=0.62026 α=2.1876 β=16.523 |
| Water level ≥16 | Poisson λ=2.8 | Burr k=0.83157 α=1.996 β=9.0268 |
| Water level ≥17 | Poisson λ=1.3 | Dagum k=0.84683 α=1.7012 β=11.927 |
| Water level ≥18 | Poisson λ=0.7 | Fatigue Life α=1.1761 β=11.19 |

Burr is a three-parameter continuous distribution with probability density function (pdf) given by

$$f(x) = \frac{\alpha k \left(\frac{x}{\beta}\right)^{\alpha-1}}{\beta\left[1+\left(\frac{x}{\beta}\right)^{\alpha}\right]^{k+1}} \tag{1}$$

and cumulative distribution function (cdf)

$$F(x) = 1 - \left(1+\left(\frac{x}{\beta}\right)^{\alpha}\right)^{-k} \tag{2}$$

Dagum, or inverse Burr, is also a three-parameter continuous distribution whose pdf and cdf can be obtained from those of Burr distribution by replacing $\frac{x}{\beta}$ by $\left(\frac{x}{\beta}\right)^{-1}$. Lastly, the two-parameter Fatigue Life, or Birnbaum-Saunders, distribution has pdf given by

$$f(x) = \frac{\sqrt{\frac{x}{\beta}}+\sqrt{\frac{\beta}{x}}}{2\alpha x}\phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}}-\sqrt{\frac{\beta}{x}}\right)\right) \tag{3}$$

and cdf

(6)

$$F(x) = \Phi\left(\frac{1}{\alpha}\left(\sqrt{\frac{x}{\beta}}-\sqrt{\frac{\beta}{x}}\right)\right) \tag{4}$$

where $\phi$ and $\Phi$ are the pdf and cdf of the standard normal distribution, respectively, and the parameters $\alpha$ and $\beta$ are positive.

Monte Carlo simulation (see Glasserman 2004 for reference) was then carried out in order to generate possible realizations of the random variables corresponding to annual frequency and duration. For each range of water level, a sample of size N=1000 from the discrete distribution of the annual frequency was generated. If the value of the random variable is $n_i$, a sample of size $n_i$ from the distribution of the random variable corresponding to duration was generated. This was done for $i = 1, 2, \ldots, 1000$. Thus, the total number of durations generated was $\sum_{i=1}^{1000} n_i$. From these values, different quantiles of the duration were obtained. The results are summarized in Table 7.

The numbers in Table 7 correspond to the maximum number of hours that water level will stay on or above the indicated lower boundary at different maximum probability levels. The first percentile gives the maximum number of hours that the water level is at least the indicated lower boundary with 1% probability. Alternatively, it gives

**Table 7.** Annual frequency of water level reaching at least a specified boundary from 2013 to 2016.

| Water Level | 1st percentile | 1st Quartile | Median | Mean | 3rd Quartile | 99th percentile |
|---|---|---|---|---|---|---|
| ≥15 m | 0.54 | 6.20 | 11.80 | 17.63 | 20.98 | 107.57 |
| ≥16 m | 0.01 | 0.94 | 3.13 | 11.78 | 9.11 | 122.86 |
| ≥17 m | 0.47 | 5.12 | 10.23 | 19.63 | 19.85 | 159.06 |
| ≥18 m | 1.25 | 5.0910 | 11.38 | 19.29 | 24.51 | 107.29 |

the minimum number of hours with 99% probability. The first quartile, median, third quartile, and 99th percentile give the maximum number of hours that the water level will be at least the indicated lower boundary with 25%, 50%, 75%, and 99% probability, respectively. For example, the number 9.11 corresponding to the row ≥16 m, column 3rd quartile means that there is 75% chance (or probability) that the water level will be at least 16 meters for at most 9.11 hours. Mathematically, if *X* represents the random variable corresponding to the number of hours that the water level is at least 16 meters and Pr denotes probability, then

$$\Pr(X \leq 9.11) = 0.75 \tag{5}$$

Alternatively, there is 25% chance that the number of hours will be at least 9.11. The probabilities 1%, 25%, 50%, 75%, and 99% are just examples; other probability levels can be calculated using the distributions in Table 6.

The numbers in the column "Mean" in Table 7 are expected values. Mathematically, if *X* is a continuous random variable with pdf *f(x),* then its mean or expected value *E(X)* is defined by

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx. \tag{6}$$

The mean is also loosely referred to as "average". In Table 7, for example, if water level reaches 17 m or higher at any time of the year, then, on the average, it will not go below 17 m for about 19.63 h (almost one day).

To complete the analysis, the annual frequency of the water level reaching at least a specified lower boundary should be considered. Recall the column corresponding to Annual Frequency in Table 6. For example, the distribution of annual frequency that the water level is at least 17 m is Poisson with parameter $\lambda$=1.3. For this, the probability of at most three occurrences is 95.7%. Thus, there is 95.7% chance that the water level will reach 17 m or higher at most thrice in a year, and for each occurrence, the water will not go below 17 m for an average of 19.63 h.

Table 8 shows the cumulative probability, called *non-exceedence probability*, that the number of times per year of water level reaching at least the indicated lower boundary will be at most the specified annual frequency. These probabilities were obtained from the distributions for the annual frequency shown in Table 6. For example, in Table 8, the non-exceedence probability of annual frequency 2 of water level >=17 is 86%. This means that the probability of at most 2 times per year of water level reaching at least 17 m is 86%. The opposite, or 1 minus the non-exceedence probability, is the *exceedence probability*. Thus, the probability of more than 2 times per year of water level reaching at least 17 m is 14%. The row MAE for N=4 means *maximum number of allowed exceedences* out of 4 data points. This will be explained and used later in the validation of the models. Likewise, the choice of values of annual frequency is for validation purposes.

**Table 8.** Non-exceedence and exceedence probabilities for annual frequencies per minimum water level.

| | Water level | | | |
|---|---|---|---|---|
| | **>=15** | **>=16** | **>=17** | **>=18** |
| Annual Frequency | 10 | 3 | 2 | 2 |
| Non-Exceedence probability | 91% | 69% | 86% | 97% |
| Exceedence probability | 9% | 31% | 14% | 3% |
| MAE for N=4 | 1 | 2 | 2 | 0 |
| Annual Frequency | 6 | 2 | 1 | 1 |
| Non-Exceedence probability | 70% | 47% | 63% | 84% |
| Exceedence probability | 30% | 53% | 37% | 16% |
| MAE for N=4 | 2 | 3 | 3 | 2 |
| Annual Frequency | 4 | | 0 | 0 |
| Non-Exceedence probability | 51% | | 27% | 50% |
| Exceedence probability | 49% | | 73% | 50% |
| MAE for N=4 | 3 | | 3 | 3 |

Table 9 shows the analogue of the numbers in Table 8 for duration in hours. This time, the value of N corresponding to the number of data points for model validation varies.

**Table 9.** Non-exceedence and exceedence probabilities for duration in hours per minimum water level.

| | Water level | | | |
|---|---|---|---|---|
| | **>=15** | **>=16** | **>=17** | **>=18** |
| **Value of N** | **24** | **9** | **4** | **3** |
| Duration (hrs) | 36 | 35 | 30 | 30 |
| Non-Exceedence probability | 90% | 90% | 85% | 81% |
| Exceedence probability | 10% | 10% | 15% | 19% |
| MAE for given N | 5 | 2 | 2 | 1 |
| Duration (hrs) | 10 | 10 | 10 | 10 |
| Non-Exceedence probability | 42% | 49% | 49% | 46% |
| Exceedence probability | 58% | 51% | 51% | 54% |
| MAE for given N | 18 | 7 | 3 | 2 |
| Duration (hrs) | 2 | 3 | 8 | 3 |
| Non-Exceedence probability | 6% | 8% | 40% | 5% |
| Exceedence probability | 94% | 92% | 60% | 95% |
| MAE for given N | 23 | 8 | 3 | 2 |

## MODEL VALIDATION

One way to check the reliability of the models presented earlier is to compare their outputs with the actual data after the period of model development. This method is commonly referred to as *backtesting*. Backtesting is a formal statistical framework that consists of verifying that actual data are in line with the projected numbers. When the model is well-calibrated, the number of observations exceeding the projected number should be in line with the specified confidence level (Jorion 2007).

To perform backtesting, hourly measurements of water level recorded at MMDA-EFCOS Sto. Nino station were once again taken, this time covering the period from 1 Jan 2013 to 31 Dec 2016. Using the same procedure as before, the length of time in hours (duration) per incident of the water level reaching at least each specified lower boundary per year was determined. The number of such incidents per year constitutes the annual frequency. The results for the duration and annual frequency are shown in Table 10.

To validate the models for the annual frequency (see Section III), the number of times that each specified projection of annual frequency for each minimum water level exceeds the MAE is counted (see Table 8). The MAE is the maximum number of allowed exceptions (MAE) at specified significance level. The level of significance is arbitrary and 5% is a usual choice. The MAE is determined by assuming that the number of exceedences, X, follows a binomial distribution with probability of success p equal to the exceedence probability (see Exceedence probabilities in Table 8) and the number of observations N being equal to the number of data points for which comparison can be made. Binomial distribution is the appropriate distribution since each observation yields exactly two possibilities: an exceedence occurs or not. Then the MAE is the highest positive integer x such that the probability that X is less than or equal to x is at most 1 - $\alpha$ = 95% that is,

$$\Pr(X \leq x; N, p) \leq 0.95 \tag{7}$$

For example, in Table 8, N = 4, and for water level at least 15 m, annual frequency 10, the exceedence probability is $p$ = 9%. Thus, the probability of exactly x exceedences is

$$\Pr(X = x; N = 4, p = 0.09) = \binom{4}{x}(0.09)^x(0.91)^{4-x} \tag{8}$$

where $\binom{4}{x} = \frac{4!}{x!(4-x)!}$. For example, the probability of at most $x$ = 1 exceedences is

$$\Pr(X = 1; N = 4, p = 0.09) = \binom{4}{1}(0.09)^1(0.91)^3 = 95.7\% \tag{9}$$

Moreover, since $\Pr(X = 0; N = 4, p = 0.09) = \binom{4}{0}(0.09)^0(0.91)^4$ = 68.57%, then the MAE is 1. That is, out of 4 observations, the number of allowed exceedences is at most 1. The other values of MAE in Table 8 are determined using the same procedure.

The results of comparison are summarized in Table 11. The Annual Frequency and Number of exceedences are based on the data in Table 9. The values of MAE are taken from Table 8. When the number of exceedences is not more than the MAE, a NO is indicated, from which it can be concluded that the model is reliable.

The procedure just described is via right-tailed hypothesis testing under the null hypothesis that p equals the exceedence probability. The test fails to reject the null hypothesis (and thus the accuracy of the annual frequency model) at $\alpha$ = 5% significance level whenever the number of exceptions does not exceed the MAE.

The same procedure is applied to the data of Duration from 2013 to 2016. The results of comparison are summarized in Table 12. Based on this, it can be concluded that the models for Duration are reliable.

**Table 10.** Duration in hours of water level reaching at least a specified boundary from 2013 to 2016 and corresponding annual frequency.

| | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|
| Duration (h) for water level >=15 | 19<br>63<br>2<br>4<br>8<br>13<br>5<br>1<br>1<br>2 | 8<br>3<br>20<br>38 | 1<br>1<br>5<br>1<br>1<br>9 | 9<br>43<br>7<br>9 |
| Annual Frequency | 10 | 4 | 6 | 4 |
| Duration (h), water level >=16 | 2<br>4<br>50 | 13<br>31 | 1<br>2 | 11<br>13 |
| Annual Frequency | 3 | 2 | 2 | 2 |
| Duration (h), water level >=17 | 12 | 7<br>26 | | 7 |
| Annual Frequency | 1 | 2 | 0 | 1 |
| Duration (h), water level >=18 | 7 | 1<br>20 | | |
| Annual Frequency | 1 | 2 | 0 | 0 |

**Table 11.** Comparison of the number of exceedences of annual frequencies from 2013 to 2016 with the MAE.

| | Water level | | | |
|---|---|---|---|---|
| | **>=15** | **>=16** | **>=17** | **>=18** |
| Annual Frequency | 10 | 3 | 2 | 2 |
| Number of exceedences | 0 | 0 | 0 | 0 |
| MAE for N=4 | 1 | 2 | 2 | 0 |
| Number of exceedences > MAE? | NO | NO | NO | NO |
| Annual Frequency | 6 | 2 | 1 | 1 |
| Number of exceedences | 1 | 1 | 1 | 1 |
| MAE for N=4 | 2 | 3 | 3 | 2 |
| Number of exceedences > MAE? | NO | NO | NO | NO |
| Annual Frequency | 4 | | 0 | 0 |
| Number of exceedences | 2 | | 3 | 2 |
| MAE for N=4 | 3 | | 3 | 3 |
| Number of exceedences > MAE? | NO | | NO | NO |

**Table 12.** Comparison of the number of exceedences of duration in hours from 2013 to 2016 with the MAE.

| | Water level | | | |
|---|---|---|---|---|
| | **>=15** | **>=16** | **>=17** | **>=18** |
| Value of N | 24 | 9 | 4 | 3 |
| Duration (hrs) | 36 | 35 | 30 | 30 |
| Number of exceedences | 3 | 1 | 0 | 0 |
| MAE for given N | 5 | 2 | 2 | 1 |
| Number of exceedences > MAE? | NO | NO | NO | NO |
| Duration (hrs) | 10 | 10 | 10 | 10 |
| Number of exceedences | 6 | 5 | 2 | 1 |
| MAE for given N | 18 | 7 | 3 | 2 |
| Number of exceedences > MAE? | NO | NO | NO | NO |
| Duration (hrs) | 2 | 3 | 8 | 3 |
| Number of exceedences | 16 | 6 | 2 | 2 |
| MAE for given N | 23 | 8 | 3 | 2 |
| Number of exceedences > MAE? | NO | NO | NO | NO |

## CONCLUSION AND RECOMMENDATIONS

In this study, ten-year historical data of water levels recorded at Sto. Niño, Marikina station of MMDA-EFCOS, from 2002 to 2012, were analysed and processed to determine the number of times per year (annual frequency) that critical levels of the Marikina River near the Sto. Niño station were reached and for how long (duration). These critical levels correspond to the four Alert Levels used by the Marikina City government to monitor the possibility of flooding and the need to evacuate residents living near the river.

From the historical data of water levels, the annual frequency and duration of reaching critical water levels were measured. The samples obtained were used to fit discrete probability distributions for the annual frequencies and continuous probability distributions for the durations. Monte Carlo simulation was then applied in order to generate possible realizations of the random variables. Summary statistics were then obtained from the simulated values.

To check the validity of the models, backtesting using historical data of water levels from 2013 to 2016 was performed. The results showed that the models obtained were reliable.

The results of this study can be used to guide the local government of Marikina in planning the needed resources in order to sufficiently respond in times of flooding incidents. It is important to note, however, that since the probability distributions and statistics were obtained from historical data, then future risk reduction measures are not yet taken into account as these could possibly change the trends for the better.

The model obtained in this study constitutes only the first step in assessing the risk of losses brought about by flooding. The next step is to determine the exposure at risk, for example, properties, lives, and livelihood that may be affected by floods. Lastly, the vulnerability of the elements exposed to flooding must be measured. These two components are recommended for further study.

## ACKNOWLEDGMENTS

This work was fully supported by a financial grant from the Philippine Higher Education Research Network (PHERNET) project of the Commission on Higher Education (CHED). The authors would also like to thank the MMDA-EFCOS for sharing their valuable data.

## REFERENCES

ABON C, DAVID C, PELLEJERA, N. 2011. Reconstructing the Tropical Storm Ketsana flood event in Marikina River, Philippines. Hydrology and Earth System Sciences 15: 1283-89.

APEL H, BISÖCHL G, THIEKEN AH, MERZ B. 2006. A Probabilistic Modelling System for Assessing Flood Risks. Natural Hazards 38(1): 79-100.

BADILLA RA. 2008. Flood Modelling in Pasig-Marikina River Basin. [MS thesis]. Enschede, Netherlands: International Institute for Geo-information Science and Earth Observation. 73p. (Available at ITC library)

BORN P, MARTIN W. 2006. Catastrophe Modeling in the Classroom. Risk Management and Insurance Review 9(2): 219-229.

CARDONA O, HURTADO J, DUQUE G, MORENO A, CHARDON A, VELASQUEZ L, PRIETO S. 2003. The Notion of Disaster Risk. Conceptual Framework for Integrated Risk Management. IDB/IDEA Program on Indicators for Disaster Risk Management, Universidad Nacional de Colombia, Manizales. 38p. http://www.manizales.unal.edu.com.

CEZAR T. 2008. Country Water Action: Flood Ready Marikina City. Case study. Asian Development Bank.

FRANCISCO JPS. 2015. Property Damage Recovery and Coping Behavior of Households Affected by Extreme Flooding Events in Marikina City, Metro Manila, Philippines. Discussion Paper No. 2015-40. Philippine Institute for Development Studies.

GLASSERMAN P. 2004. Monte Carlo Methods in Financial Engineering: Springer. 599p.

GROSSI P, KUNREUTHER H. 2005. Catastrophe Modeling: A New Approach to Managing Risk. Boston: Springer. 266p.

IGLESIAS GR. 2008. Flood Disaster Mitigation and River Rehabilitation in Marikina City. Case study, Safer Cities 22: 1-8. Asian Disaster Preparedness Center.

JORION P. 2007. Value at Risk: The New Benchmarrk for Managing Financial Risk, 3rd Edition: McGraw-Hill. 602p.

KUNDU PM, MATHIVHA FI, NKUNA TR, ODIYO JO, SINGO LR. 2012. Flood Frequency Analysis of Annual Maximum Stream Flows for Luvuvhu River Catchment, Limpopo Province, South Africa. Conference paper in the 16th South America National Committee at the International Association of Hydrological Scientists (SANCIAHS) National Hydrology Symposium; 2012 October 1-3; Pretoria, South Africa: SANCIAHS.

MECHLER R and The Risk to Resilience Study Team. 2008. The Cost-Benefit Analysis Methodology. From Risk to Resilience Working Paper No. 1. ProVention Consortium, Institute for Social and Environmental Transition and Institute for Social and Environmental Transition-Nepal.

MECHLER R. 2005. Cost-benefit Analysis of Natural Disaster Risk Management in Developing Countries. Disaster Risk Management in Development Cooperation Sector Project. Deutsche Gellenschaft für Technische Zusammenarcht (GTZ) GmbH.

MICHEL-KERJAN E, HOCHRAINER-STIGLER S, KUNREUTHER H, LINNEROOTH-BAYER J, MECHLER R, , MUIR-WOOD R, RANGER N, VAZIR P, YOUNG M. 2013. Catastrophe Risk Models for Evaluating Disaster Risk Reduction Investments in Developing Countries. Risk Analysis 33(6): 984-998.

[NOAH] National Operational Assessment of Hazards. Ondoy (2009), Floods in Marikina City, Metro Manila, Disaster Timeline. Retrieved from http://blog.noah.dost.gov.ph.

PANJER H. 2006. Operational Risks, Modeling Analytics. John Wiley & Sons.

SATO T, NAKASU T. 2011. 2009 Typhoon Ondoy Flood Disasters in Metro Manila. Natural Disaster Research Report of the National Research Institute for Earth Science and Disaster Prevention 45: 63-74.

YU CZ, SAJOR EE. 2008. Urban Rehabilitation: A Case Study in Marikina City. Case study. Water Environment Partnership in Asia.